Low-frequency Active Target Characterization Using Hidden Markov Models and Classifiers

David H. Kil
and
Frances B. Shin,

Lockheed Sanders, Inc.
Advanced Engineering & Technology Division
Signal Processing Center of Technology
PTP2-A001
P.O. Box 868
Nashua, NH 03061-0868

and

J. Robert Fricke

Department of Ocean Engineering
Massachusetts Institute of Technology
77 Massachusetts Avenue, Room 5-218
Cambridge, MA 02139

for
Dr. Kam W. Ng
Contract No. N00014-91-C-0125
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217-5660

19951031 040

19951031 040

Accession For
NTIS CRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification .....
By per attached
Distribution /
Availability Codes
Dist | Avail and / or Special
A-1

Contents

List of Figures

Abstract

We investigate various projection spaces and extract key parameters or features from each space to characterize low-frequency active (LFA) target returns in a low-dimensional space. The projection spaces encompass (1) time-embedded phase map, (2) segmented matched filter output, (3) various time-frequency distribution functions, such as Reduced Interference Distribution, to capture time-varying echo signatures, and (4) principal component inversion for signal cleaning and characterization. We utilize both dynamic and static features and parameterize them with a hybrid classification methodology consisting of hidden Markov models, classifiers, and data fusion. This clue identification and evaluation process is complemented by concurrent work on target physics to enhance our understanding of the target echo formation process. As a function of target aspect, we can observe (1) back scatter dominated by axial n=0 modes propagating back and forth along the length of the shell, (2) direct scatter from shell discontinuities, (3) helical or creeping waves from phase matching between the acoustic waves and membrane waves (both shear and compressional), and (4) the "array response" of the shell, with coherent superposition of elemental scattering sites along the shell leading to a peak response near broadside. As a function of target structures (the empty shell and the ribbed/complex shells), we see considerable complexity brought about by multiple reflections of the membrane waves between the rings. We show the merit of fusing parameters estimated from these projection spaces in characterizing LFA target returns using the MIT/NRL scaled model data. Our hybrid classifiers outperform the matched filter-based recognizer by an average of 5 to 25%. This improvement can be attributed to a combination of good features that maximize inter-class discrimination and appropriate classifier topologies that exploit the underlying multi-dimensional feature probability density function.

1 INTRODUCTION

Low-frequency active target echo characterization is of considerable interest from the perspective of target physics and signal processing because of complex, time-varying echo structures. In order to facilitate our understanding of target physics, the Massachusetts Institute of Technology (MIT) in collaboration with the Naval Research Laboratory (NRL) conducted a scaled model experiment in which monostatic and bistatic returns were recorded for three cylinder types along the 360 degree azimuthal sector. For $2 < ka < 10$, where k and a refer to the wavenumber and the radius of the cylinder, respectively, backscattered returns from finite cylinders consist of early and mid-to-late returns. The early returns are dominated by backscatter from target discontinuities for angles away from broadside and by the target "array response" at broadside generated by coherent superposition of elemental scattering along the length of the shell. The mid-to-late returns are dominated by supersonic helical or creeping waves produced by phase matching between the acoustic and membrane waves (both shear and compressional), and by slow flexural waves interacting with discontinuities in the shell. The echo return structure varies as a function of target type, aspect, ka, and time.

In order to take advantage of the variable echo structure, we investigate classifier topologies appropriate for echo characterization. Hidden Markov model (HMM) (1) is one of the most popular techniques to characterize time-varying patterns . The HMM models temporal variation in the feature space with a finite number of states, state transition probability matrices, observation probabilities for each state, and initial state occupancy probabilities. Features refer to parameters that capture essential target attributes useful for target characterization and discrimination. One drawback of the HMM-based recognition paradigm is that although the HMM maximizes the class likelihood ratio, it does not address the issue of inter-class discrimination, which is key to achieving good target recognition performance (2). On the other hand, both conventional and neural net classifiers emphasize discrimination when coupled with an appropriate feature optimization and rank order algorithm. However, they generally lack a mechanism to explicitly accommodate temporal variations.

Therefore, we develop a *reconfigurable classifier architecture* that combines the advantages of the HMM and classifiers. Our integrated target characterization paradigm can be succinctly described by:

1. low-dimensional data projection for feature extraction,
2. ranking of features and time segments in terms of their contribution to overall classification,
3. selection of an appropriate classifier architecture that is best mapped to the underlying multi-dimensional feature probability density function, and
4. recognition performance analysis in terms of rank order curves, confusion matrices, and classification receiver operating characteristic (ROC) curves.

Figure 1 depicts the integrated classification methodology that maximizes both discrimination and likelihood.

2 TARGET PHYSICS

In this paper we work exclusively with laboratory scale model backscatter data for finite cylindrical shells. To facilitate interpretation of features extracted by the low-dimensional classifiers, it is useful to have an underlying understanding of the target physics. In particular, features in the data can be attributed to specific physical mechanisms, which are linked to both the target type and the ensonification aspect. The next section briefly reviews the experimental design followed by a section reviewing the dominant signal physics and echo formation.

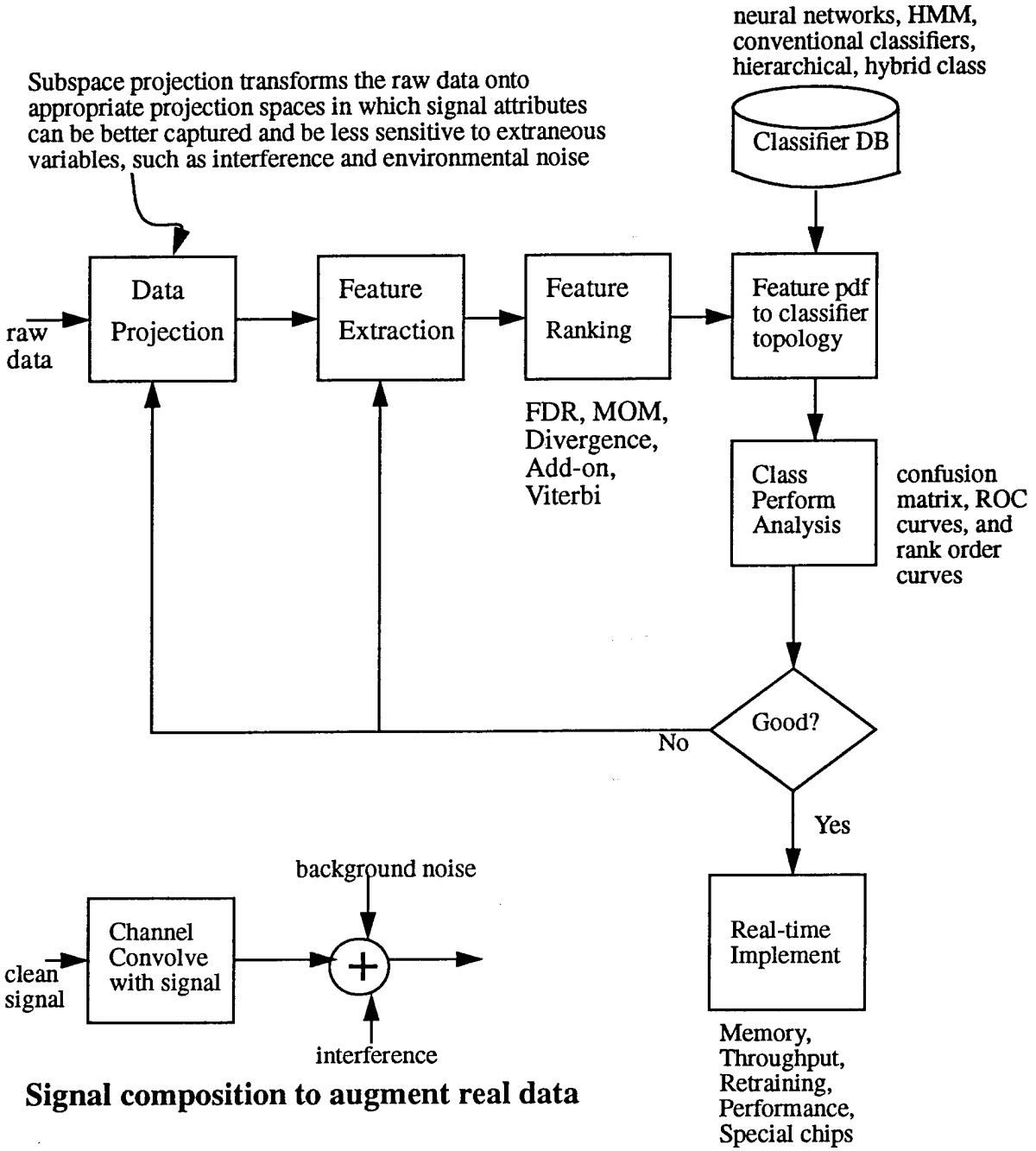


Figure 1: Our integrated classification paradigm consists of low-dimensional projection, feature optimization, matching the classifier topology to the underlying feature distribution, and performance analysis.

2.1 Experiment

The experimental data used in this paper was collected at NRL in collaboration with the ONR sponsored structural acoustics research program at MIT. Three finite cylinders were fabricated at varying levels of complexity. The external appearance of all the models is identical. They are 0.862 m in length with a diameter of 0.111 m. The skin is Ni-200 nickel with thickness of $0.532 \text{x} 10^{-3}$ m, which yields a 100:1 shell thickness to radius ratio. The shells differ in their internal configuration. The simplest has no internal structure, only the shell plating is present. The next level of complexity involves the placement of four unequally spaced deep-rings with an aggregate mass equal to that of the shell. The highest level of complexity is a shell with four suspended masses isolated from the rings with rubber bulkheads at each ring location and four delrin rods connecting the masses running the length of the shell. This model is not intended to mimic a full scale target, but rather to produce acoustic complexity comparable to that of a full scale target. The three models are known respectively as the "empty", "ribbed", and "complex" models. For the purposes of this paper scatter from the three models may be classified into two categories: (1) scatter from the empty model and (2) scatter from the ribbed and complex model. As may be inferred from this statement, the ribs play a dominant role in scattering process of the complex shell. Both monostatic and bistatic measurements were taken for each shell, but only the monostatic measurements are discussed in this paper. Further, since the ribbed and complex cylinders are so similar, of the two only the ribbed data will be discussed in comparison with the empty shell data.

The models were placed in an acoustic underwater test facility at NRL and ensonified by a plane wave source as shown in Figure 2. For the monostatic measurements the source array and single receiver hydrophone remained stationary. The models were rotated through a range of 360° in 1° steps. Since the problem is approximately quadrant symmetric only azimuth angles from 0 degrees (bow) to 90 degrees (beam) are considered here. This range shows the salient properties of scattering from finite cylinders. At each angle 100 pings were averaged to compute the backscattered return from the target. The source waveform was a wide-band pulse with useful energy in the range of 10-50 kHz, which corresponds to $2 < ka < 10$. To minimize the effects of clutter from the measurement tank a clutter subtraction process was used to clean the data. This process consisted of measuring the tank response with no target in place and subtracting this response from that measured with the target present. Details of the acquisition and initial data analysis are discussed by Corrado (3).

This paper presents an alternate analysis approach to the same data set, as is described below. To understand the analysis, however, we must first consider some of the fundamental physical processes associated with the target scattering. Much of this understanding is due to Corrado's work.

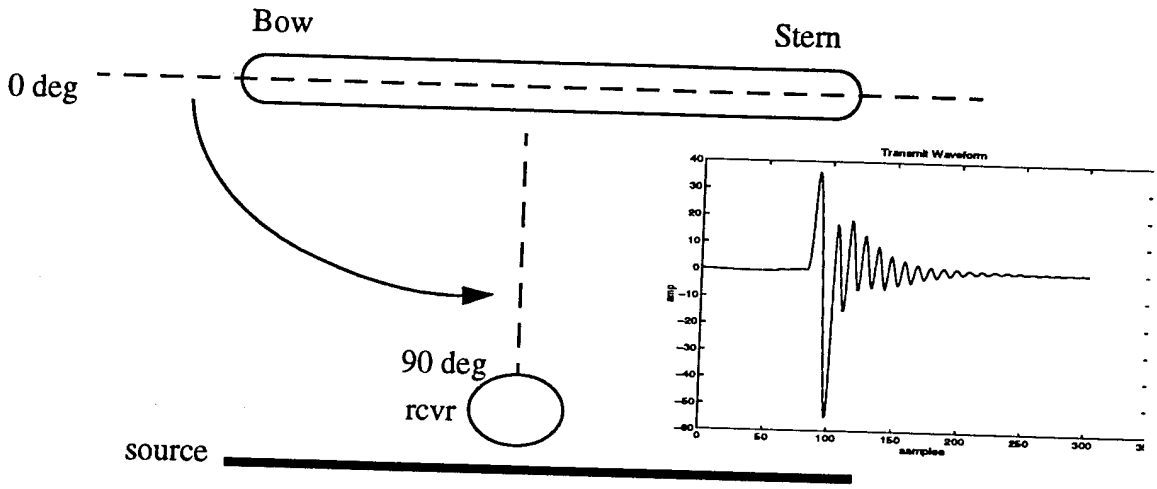


Figure 2: Experimental geometry and the transmit waveform.

2.2 Physical Interpretation

The data may be divided into two major classification groupings as shown in Figure 3.

Namely,

1. time zones: early, mid, and late
2. azimuthal zones:
   (a) $\pm 20^o$ bow/stern
   (b) $20^o < \theta < 60^o$ low return sector
   (c) $60^o < \theta < 85^o$ helical wave sector
   (d) $85^o < \theta < 90^o$ beam sector

Of these 12 regions only six are of importance for this paper: all the early time regions, the mid-time membrane zone region, and the late-time bow zone region. Each of these regions is dominated by one of several physical processes, which are, in large part, individually determined by a given wave type: compressional, shear, and flexural. For the ribbed case, the boundary between the low-return zone and the helical zone is not distinct due to scattering from the rings.

In order to facilitate understanding of target echo formation process, we used the Reduced Interference Distribution (RID) (4) to project raw time series onto a high

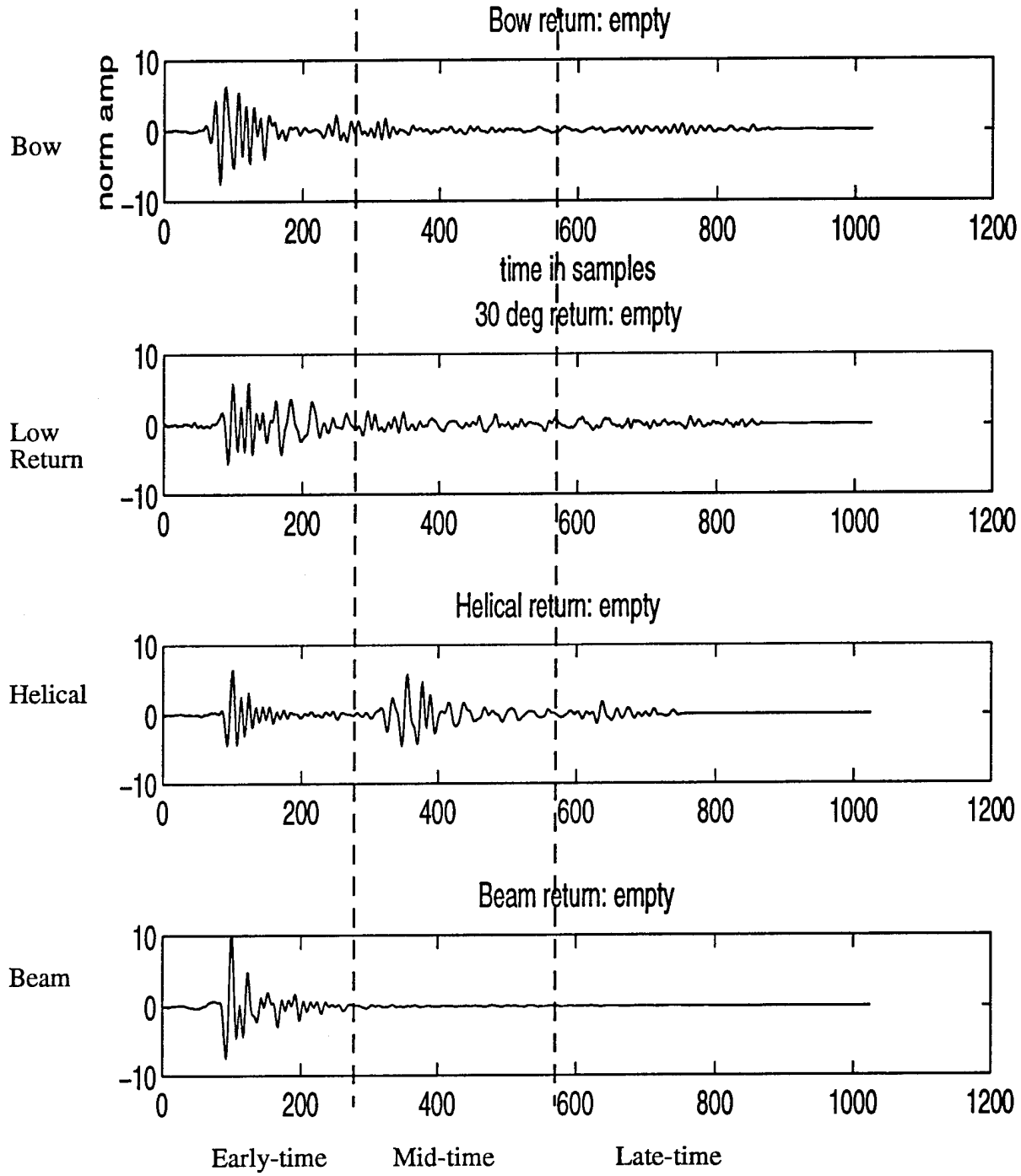


Figure 3: Echo groupings as a function of time and aspect for LFA return characterization. The sampling period is 4 microsec.

resolution time-frequency map as shown in Figure 4. The RID is particularly useful when the signal consists of a number of overlapping components in time and frequency which can cause "cross-term" interferences.

Let us first consider the four major azimuthal divisions, which are approximately quadrant symmetric. At bow incidence ($\pm 20^o$) n=0 modes dominate. The ability to excite higher circumferential modes is low due to the plane wave excitation which produces axisymmetric pressures around the shell. These n=0 modes propagate back and forth down the hull as though it were a transmission line (5) scattering from discontinuities as it goes. Next consider the helical zone, which is so called because here the incident acoustic wave is phase, or trace, matched to both compressional and shear helical waves. These waves are supersonic and have an axial component and a circumferential component, which results in a winding helical propagation path down the shell. The circumferential mode number for helical waves is greater than zero due to the unsymmetric forcing of the shell by the incident wave at these angles. Now, the low-return zone is simply an area where the axisymmetric forcing of n=0 modes is weak and trace matching to helical wave cannot take place due to geometric considerations. The main return here is direct scatter from the target discontinuities. Finally, the beam zone is dominated by direct backscatter from the entire shell length due to the coherent superposition of elemental scattering centers. These returns may be thought of as an "array response" for the shell, and the monostatic return is simply the peak response of the beam pattern near broadside.

Now consider the temporal divisions of the data. At early time the dominant backscatter is due to direct acoustic scatter from discontinuities in the shell including endcaps, slope discontinuities, and rings. At each discontinuity every wave type scatters into every other wave type to varying degrees. As an example, at bow incidence the earliest return is direct scatter off the endcap followed about 200 μs later by a return, which derives from direct excitation of a flexural wave in the endcap that then backscatters an acoustic wave when it encounters the slope discontinuity between the endcap and the cylindrical shell (6). At mid-time, we continue to see direct scatter from the ribs and far endcap over a broad range of angles. The more interesting feature, however, is the formation of the helical waves seen in a $\pm 30^o$ sector with respect to beam for the empty shell and $\pm 45^o$ for the ribbed shell. The difference between the shells is due to multiple scattering of the helical wave between the rings as it winds down the shell. All late time events are almost certainly due to flexural waves propagating slowly down the shell scattering into acoustic energy at discontinuities. Since the flexural wave is dispersive, we expect to see frequency down swept chirps for the late returns. For the empty shell, which produces particularly simple signal structure, the down-chirps are relatively distinct.

Comparing the empty and ring stiffened shell responses one sees considerable complexity brought about by multiple reflections of the helical waves between the rings. Early time response is due to compressional and shear waves in the skin scattering from the rings and the endcaps. Late time response derives from flexural waves slowly lumbering their way down the shell and scattering through mode conversion

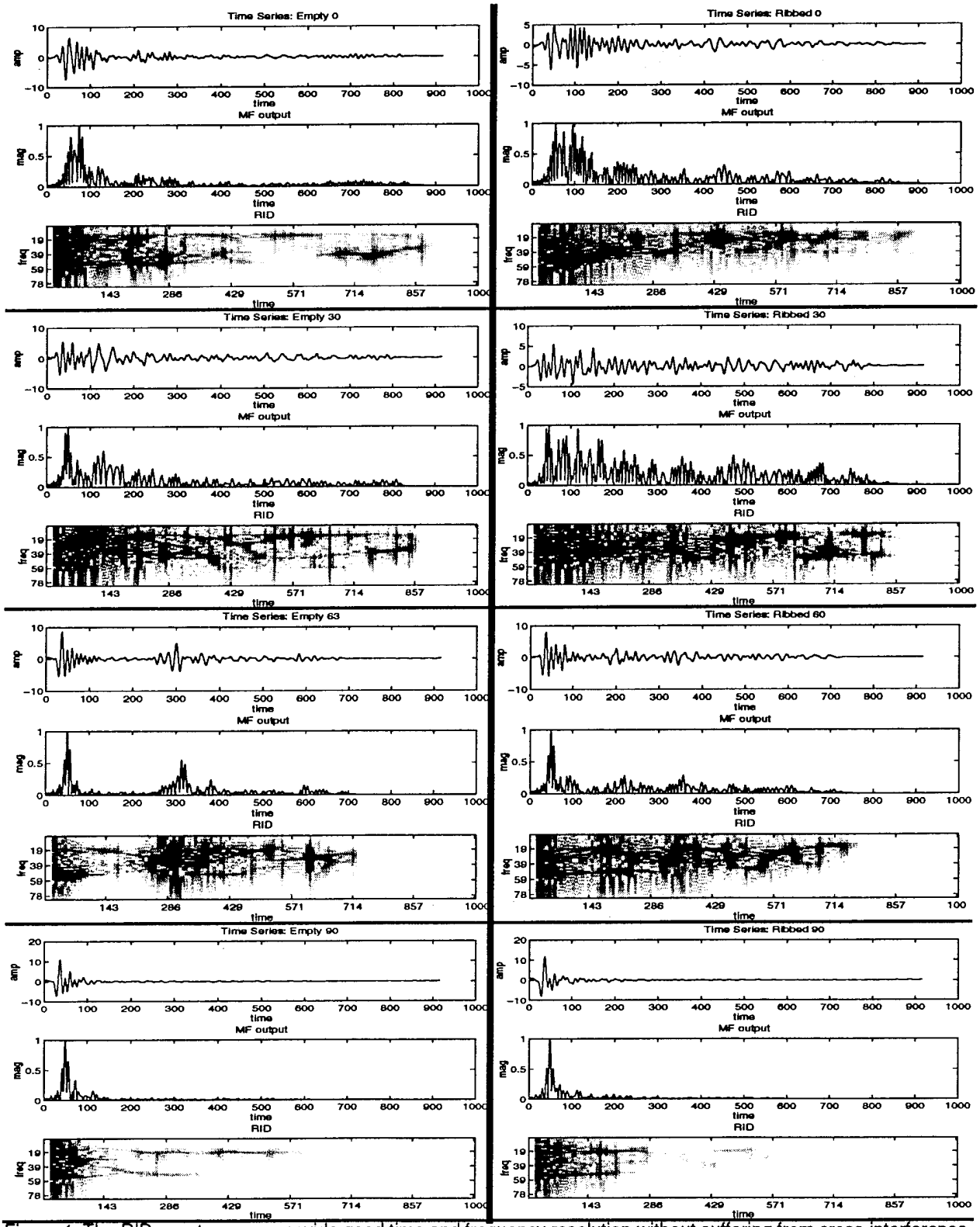


Figure 4: The RID spectrograms provide good time and frequency resolution without suffering from cross-interference terms. Time-series and matched filter outputs are also shown for comparison. The left plots show empty target echos with 0, 30, 60, and 90 deg aspect groups from top to bottom while the right plots show ribbed target returns in the same aspect group order. The time axis is in samples where the sampling period is 4 microsec. The frequency axis is in kHz covering ka of 0 to 17.

at the rings and endcaps. At bow incidence, one can see the result of multiple reflections bouncing back and forth in the first bay of the shell. In contrast, the empty shell response is dominated by scattering from the endcaps with a longer delay time between bounces. The observed acoustic energy derives both from direct scatter of compressional and shear waves and from mode conversion of flexural waves. For the ribbed shell the late time events between 30^o and 60^o are especially hard to interpret, since their nature is deterministic but very complicated in terms of multipath and mode conversion at discontinuities. The 0^o and 90^o cases are end members, and "relatively" simple.

3 PROJECTION SPACE INVESTIGATION

The conceptual framework of low-dimensional projection is based on data compression which uses as few bits of information as possible to convey the entire message with the lowest possible bit error rate. One simple example is to use an oversampled Gabor representation to characterize exponentially damped sinusoids. Advantages of extracting features from the low-dimensional projection space encompass (1) robustness to extraneous variables such as noise due to energy compaction, (2) computational efficiency in classification due to inherent data compression, and (3) facilitation of feature discriminant analysis.

After our thorough investigation of target physics, we explore the following four projection spaces for target characterization. We select them because they provide an accurate time "snapshot" of spectral contents. Furthermore, they perform adaptive smoothing and/or filtering to remove as much out-of-band noise as possible. In short, these projection spaces provide a temporal map where appropriate clues or features can be extracted at each time snapshot after noise filtering. The goal at this stage is to extract as many pertinent features as possible from each projection space for later feature optimization and fusion.

3.1 Reduced Interference Distribution (RID) Spectrogram

The RID allows us to construct a "time-frequency acoustic signature" of an echo. Instead of using the entire RID spectrogram, we implement an innovative image compression algorithm to extract features from the RID spectrogram. The image compression algorithm consists of two-dimensional transform for further coefficient compaction, transform coefficient encoding, encoded transform coefficient compression with the singular value decomposition (SVD) to overcome the "curse of dimensionality", vector quantization (VQ), and entropy or arithmetic encoding of the VQ codebook indexes (14). We perform the coefficient compression by using the signal subspace eigenvectors where we use the minimum description length (MDL) criterion (7) to determine the rank of the covariance matrix. Again a combination of transform and SVD-based subspace filtering results in data compaction and improved SNR for robust target characterization.

3.2 Segmented Matched Filter Output

We divide the matched filter output into a number of equal time segments. From each time segment, we extract shape and amplitude statistics (i.e., mean, standard deviation, skewness, and kurtosis). We also extract the same statistical parameters from the difference between the matched filter output and the raw energy detector output. This is done to evaluate the discrepancy between the correlated and uncorrelated components of the return energy with the transmit waveform, especially in the late arrival segment.

3.3 Principal Component Inversion (PCI) Output

The PCI methodology works as an adaptive Wiener filter in that it estimates the time-varying signal structure and utilizes the SVD to separate the data into signal subspace and alternate subspace components (8). From the signal subspace component, we can estimate the clean signal structure by using diagonal averaging. From each PCI output, we extract center frequency, bandwidth, linear predictive coding (LPC) coefficients, cepstral coefficients, delta-cepstral coefficients, state transition parameters, singular value distribution, and low-rank dimension. Figure 5 illustrates the noise reduction performance improvement with PCI in comparison to a conventional filtering matched to the transmit pulse bandwidth.

3.4 Compressed Phase Map with the SVD

In nonlinear dynamical modeling of chaotic structures, the time-embedded representation or phase map is often used. An embedding dimension of two is sufficient to characterize the Henon noise whose dynamics are governed by a second order differential equation (9). However, for the MIT/NRL tank data, we do not observe such a low-order, deterministic structure in the phase map. Therefore, we use an embedding dimension of 32, but perform data compression using the SVD so that only the first eight principal components are used for data analysis.

4 INTEGRATED TARGET CHARACTERIZATION PARADIGM

After feature extraction, we perform thorough feature discriminant analysis followed by target characterization performance assessment. For this analysis, we generate eight classes: two target types (empty and ribbed) and four aspect groups (near 0^o, 30^o, 60^o, and 90^o). Our goal is to find a good feature subset and an appropriate classifier topology matched to the underlying good feature distribution.

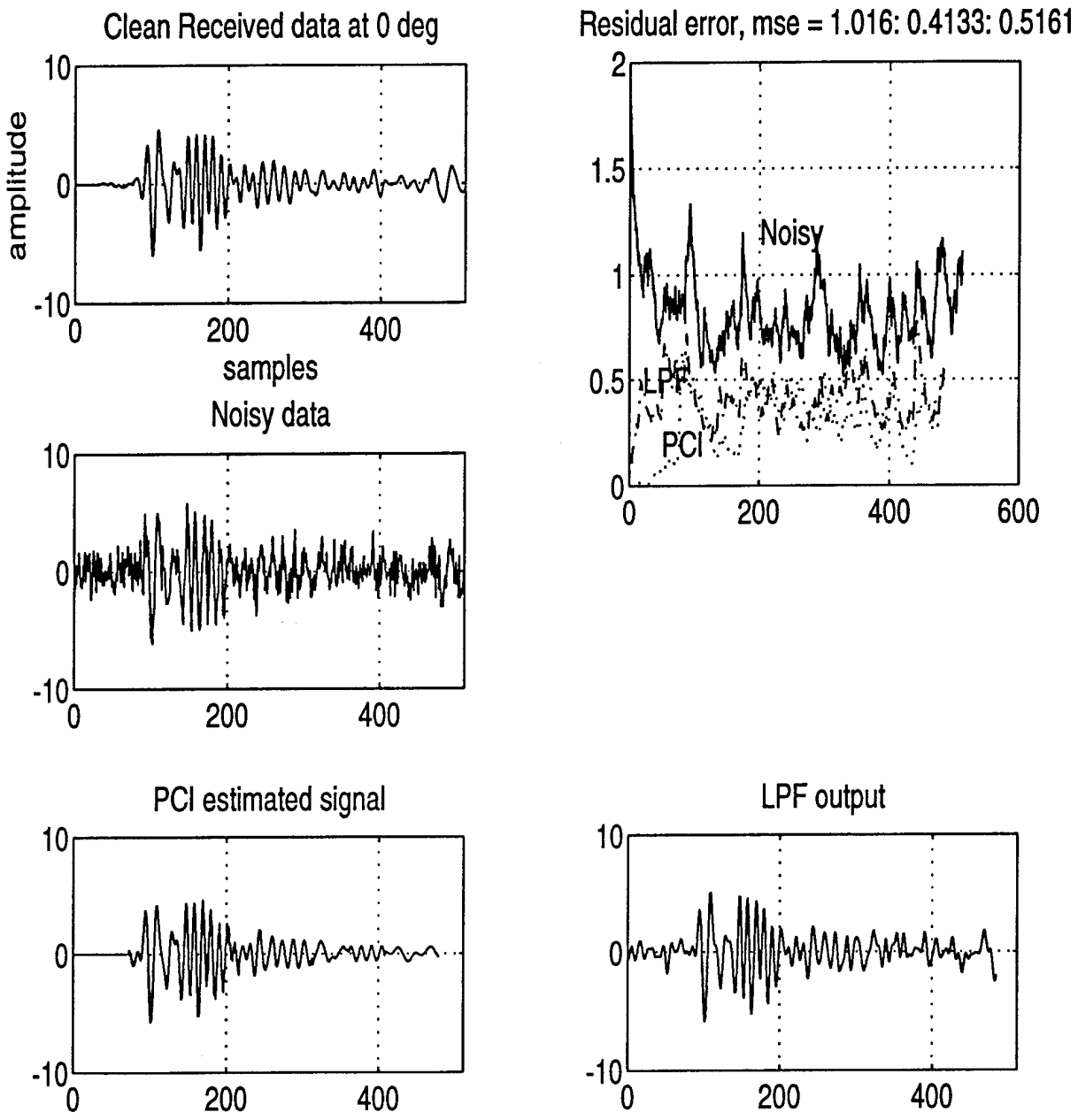


Figure 5: PCI performs time-adaptive filtering and yields better noise rejection performance than the low-pass filtering.

4.1 Feature Discriminant Analysis

In order to maximize class separability for robust recognition performance, we evaluate features in terms of class discrimination and the degree of feature correlation. Feature analysis algorithms consist of Fisher's discriminant ratio (FDR), Procrustes angle, multi-modal overlap measure (MOM), divergence, add-on/knock-out, Viterbi, and projection-based discriminant ratio tests. While one dimensional feature optimization algorithms, such as FDR, MOM, divergence, and Procrustes angle, are fast, they fail to take feature correlation into consideration. Further, when single dimensional feature distributions exhibit a high degree of overlap among classes, the one dimensional feature optimization algorithms yield inferior performance to that of multi-dimensional feature optimization algorithms. Unfortunately, the add-on or Viterbi algorithms are computationally expensive, especially for a large number of training tokens. In an attempt to combine the strengths of single and multi-dimensional feature ranking algorithms, projection-based algorithms perform feature compression followed by discriminant analysis in the compressed or reduced feature space.

4.1.1 *Linear Fisher's Discriminant Ratio (FDR)*

The FDR is a statistical rank order method which determines a feature priority by computing a detection index, $\frac{\Delta\mu}{\sigma}$, between classes where $\Delta\mu$ and σ refer to the mean difference between any pair of classes and feature standard deviation, respectively. Generally, FDR is ideal for unimodally distributed, Gaussian features and can be computed as follows:

$$FDR1(p) = \sum_{i=1}^{K} \sum_{i \neq j}^{K} \frac{\left(\mu_p^i - \mu_p^j\right)^2}{\left(\sigma_p^i\right)^2 + \left(\sigma_p^j\right)^2}, \tag{1}$$

$$FDR2(p) = \frac{\sum_{i=1}^{K} \sum_{j=1}^{K} \left(\mu_p^i - \mu_p^j\right)^2}{\sum_{i=1}^{K} \left(\sigma_p^i\right)^2}, \tag{2}$$

where

$$\begin{aligned} K &= \text{the number of classes,} \\ p &= \text{feature index,} \\ \mu_p^i &= \text{mean of the } p^{th} \text{ feature for the } i^{th} \text{ class, and} \\ \sigma_p^i &= \text{standard deviation of the } p^{th} \text{ feature for the } i^{th} \text{ class.} \end{aligned} \tag{3}$$

The main difference between FDR1 and FDR2 is that FDR1 tends to emphasize separation between any two classes while FDR2 averages over all the classes. Therefore,

FDR2 is more appropriate for rank ordering features with more than two classes.

For problems involving a large number of classes, instead of summing detection indexes over all the classes, we can use the worst-case detection index to rank order individual features for more robust recognition performance. This concept of making the worst-case performance as favorable as possible is the backbone of the minimax algorithm.

4.1.2 *Procrustes Angle*

Procrustes angle (10,11) is closely related to the least squares approximation and measures the relationship between two given subspaces. Since the vector subspace defined by the eigenvectors corresponding to the significant eigenvalues of the Fisher covariance matrix is optimal in the least squares sense, it is intuitive that the angle between the k^{th} feature and its orthogonal projection onto the Fisher projection subspace should be small for good features, and large for less useful ones. This formulation is conceptually similar to linear Fisher's discriminant analysis.

4.1.3 *Multi-modal Overlap Measure*

MOM is appropriate for features which exhibit multi-modal and non-Gaussian probability density functions. Feature rank is determined by integrating the area of overlap between class pdfs. As expected, features with the least degree of overlap are assigned the highest ranks.

Another discriminant measure based on the estimated feature pdfs is often referred to as divergence and can be computed as follows (12):

$$D_{ij}(k) = I_k(i,j) + I_k(j,i) \tag{4}$$

where

$$\begin{aligned} I_k(i,j) &= \int_x p_i^k(x) \ln\left(\frac{p_i^k(x)}{p_j^k(x)}\right) dx, \qquad (5) \\ k &= \text{feature index}, \\ p_i^k(x) &= \text{pdf of class i for feature k, and} \\ p_j^k(x) &= \text{pdf of class j for feature k.} \end{aligned}$$

For multi-class problems, following the minimax approach, we can attempt to maximize the worst-case performance by rank ordering features based on the minimum value of $D_{ij}(k)$ over i and j instead of summing $D_{ij}(k)$ over i and j (i.e., $D(k) = \sum_{i<j} \sum_j D_{ij}(k)$) (2). Although a little pessimistic in its philosophy, this minimax or "maximin" approach can yield robust recognition performance under certain situations.

4.1.4 *Feature Optimization in Multiple Feature Dimensions*

For difficult problems with very complex class boundary functions and a substantial amount of overlap in the single dimensional feature space, it is advantageous to perform multi-dimensional feature optimization. Conceptually similar to its original use in convolutional encoding and decoding in communication, Viterbi algorithm (13) or dynamic programming considers many subsets in parallel to find the best M feature subset out of N candidate features ($M \leq N$). If the performance measure increases monotonically as a function of the feature subset size and the performance at any stage is a function of the previous feature subset and the current feature (i.e., Markov property), then this process will result in the optimum M feature subset.

The Viterbi rank order procedure is summarized below.

1. Evaluate the performance of N subsets, each consisting of one feature.
2. For each subset of one feature, append one of the remaining N-1 features, evaluate the performance of N-1 two-feature subsets, and select the two-feature subset that yields the best performance.
3. Now for each subset of two features, append one of the remaining N-2 features and select the three-feature subset that yields the best performance. Repeat the same procedure until each subset contains M features or performance degradation occurs.
4. Select the path that yields the best performance. Features that fall into the optimal path constitute the feature subset to be used for classification.

Although dynamic programming is more computationally tractable than the exhaustive search method, it is still very time-consuming. As a consequence, for all practical problems, we resort to suboptimal add-on or knock-out algorithms to find a "reasonably" good feature subset. The only difference between add-on/knock-out and Viterbi is that the former considers only one best path at any stage, thereby saving computational loading by a factor of N. Based on our extensive classification experiences, the performance difference between add-on and Viterbi is approximately 0 to 4 %.

In general, multi-dimensional feature optimization tends to yield optimistic recognition performance. The optimal feature subset composition is likely to change from run to run, provided that random cross validation is performed to independently assess the recognition performance. This means that the feature subset composition may not remain fixed during random cross validation. That is, the recognition performance averaged over multiple random runs may be based on different feature subsets. Therefore, the classification performance based on the Viterbi and add-on algorithms provides the theoretically attainable upper bound (i.e., similar to Cramer-Rao Lower Bounds).

4.2 Classifier Architecture

Conventional and neural network classifiers have been used extensively in pattern recognition. They can be divided into parametric, non-parametric, and boundary decision classifiers. Parametric classifiers, such as a multi-variate Gaussian classifier (MVG), makes a certain statistical assumption regarding the underlying feature pdf, resulting in recognition performance that depends on the goodness of the statistical fit. Non-parametric classifiers, such as a binary tree classifier, make no such assumption and generally require a large number of training tokens. Boundary decision classifiers, such as a backpropagation neural net (BPN), attempt to find a class boundary function that best separates classes based on some error criteria and typically suffer from long training time and possible convergence to one of the local minima. As a result of these differences among classifiers, it is crucial that we perform thorough feature discriminant analysis and match the classifier architecture to the underlying feature pdf.

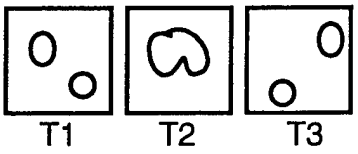
HMMs are popular in modeling temporal variability. Feature tokens are computed at an appropriate frame rate. The HMM models feature variation in time by assigning tokens with similar statistical characteristics to a common state. Within each state, the observation probability of all the tokens is usually based on a Gaussian mixture model with a diagonal or full covariance matrix. In speech modeling, the left-to-right state model is widely used to closely follow speech articulation. In nonlinear dynamical modeling, each state typically represents a cluster in the time-embedded phase map. The HMM model parameters, consisting of initial state occupancy, state transition, and observation probabilities, are estimated using either segmental k-means or forward-backward algorithms. Since the transition matrix is full (i.e., no zero elements), we will denote such HMMs "ergodic".

Although the HMM is quite useful in statistical characterization of time-varying dynamic patterns, one potential drawback of the HMM is that it tends to maximize the likelihood of the correct class, but does not suppress the likelihood of other incorrect classes. That is, the HMM does not address an important issue of discrimination and robustness which are key to achieving good classification performance. Therefore, it makes sense to emphasize class discrimination during training. In order to design a classifier topology that combines the merits of classifiers with good class discrimination and HMMs with good temporal variability characterization, we present the three reconfigurable classifier architectures for the LFA target echo characterization as depicted in Figure 6.

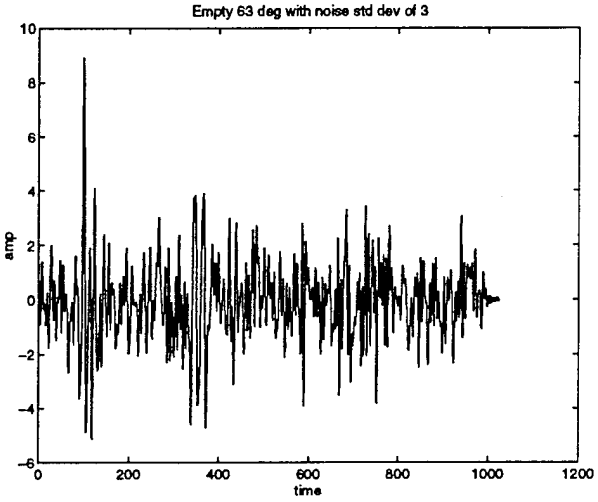
4.2.1 *Left-to-right HMM*

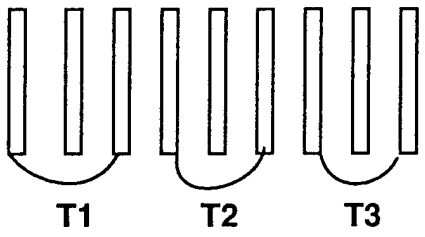
This is the classic HMM used in speech recognition. The one exception in our implementation is that instead of using all the features in the likelihood ratio computation, we use only those that provide good class discrimination. Feature ranking is performed as a function of time (i.e., state in this case). Since the HMM generally uses the same feature subset for the log-likelihood ratio (LLR) score computation,

Time-varying MMC



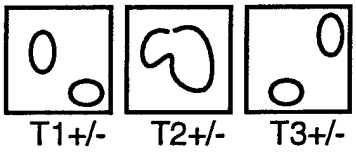
1. state = fixed or var time segment
2. separate feature ranking for each state
3. flexible to use any classifier for each state
4. emphasis on class separation
5. more robust than HMM





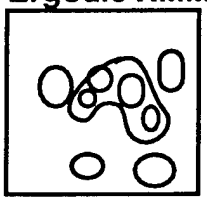
Feature vectors computed at an appropriate frame rate

Left-to-right model HMM



1. state = slightly variable time segment
2. usually same features throughout
3. good for dealing with speech and speaker variabilities
4. usually uses the GMM to characterize obs. prob. associated with each state
5. emphasis on maximizing likelihood with SKM/B-W

Ergodic HMM



1. state = distinct cluster
2. more robust than L-to-R HMM
3. derived from nonlinear dynamical modeling of low-order deterministic signals
4. flexibility in state transition
5. emphasis on maximizing likelihood with SKM/B-W

Figure 6: The three classifier topologies attempt to maximize both discrimination and likelihood by taking advantage of feature optimization and by accommodating temporal variability.

feature ranking is done globally over the entire observation period. Furthermore, we identify and exclude time segments that add to confusion. Within each state, we use the Gaussian mixture model for characterizing observation probabilities.

4.2.2 *Ergodic HMM*

After feature optimization and ranking, we populate the multi-dimensional feature space with all the feature tokens from each class. Now we perform the VQ to find distinct clusters or states. We can trade-off the number of states versus the complexity of modeling observation probability for each state. Unlike the left-to-right HMM, the transition probability matrix of this HMM is usually full.

4.2.3 *Temporally Adaptive Classifier*

The structure of this classifier is similar to the left-to-right HMM with the following two exceptions:

1. Features are optimized separately for each state (i.e., local feature optimization). This approach allows the maximum flexibility in the classifier and feature architecture.
2. Any classifier can be assigned to any state as long as the selected classifier provides the best fit to the underlying feature pdf.

5 MIT/NRL SCALED MODEL DATA ANALYSIS RESULTS

Due to the symmetric nature of the target echo structure, we focus our characterization efforts on 0 to 90 degree aspect for the empty and ribbed targets. We subdivided the quadrant into near 0^o ($\pm 5^o$), near 30^o ($\pm 5^o$), near 60^o ($\pm 5^o$), and near 90^o ($\pm 5^o$). Therefore, the task of a recognizer is to determine the aspect and target type of an echo corrupted in white Gaussian noise.

The training data consists of a clean data with the signal-to-reverberation ratio of at least 10 dB. We normalize the clean data so that its mean and standard deviation are 0 and 1, respectively. For testing at various SNR's, we corrupt the clean data with independent white Gaussian noise.

5.1 Recognition Performance With Matched Filters

At first, we pose the following question: what if we use the received target echo as a matched filter? By contrast, the matched filter projection space for feature extraction utilizes the transmit waveform, not the received waveform. In essence,

classifiers use templates or features of known classes for pattern matching. Figure 7 shows confusion matrices along with the overall, and individual class recognition performances as a function of the noise standard deviation.

Confusion matrix, as the name implies, is a measure of how classifiers respond given an input signal with unknown identity. It can be generated by inputing a large number of test feature vectors and by grading the classifier outputs against the known ground truth. The correct recognition values can be read from the diagonal elements of the confusion matrix and off-diagonal elements indicate the degree of confusion with incorrect classes. Most confusion occurs for the empty 60^o and empty/ribbed 90^o classes because of a good deal of variability as a function of aspect and the lack of any late return structure, respectively. Our next task is to evaluate the three classifier candidates to determine if they can outperform the matched filter by a judicious combination of good features and appropriate classifier architecture.

5.2 Multi-dimensional Feature Distribution

After feature extraction and ranking, we look into the multi-dimensional feature pdf to derive the matching classifier architecture. Figure 8 illustrates the compressed feature scatter plot as a function of time. Note the highly non-Gaussian and multi-modal feature distribution which will make the MVG a poor choice for this problem. Moreover, there is a considerable amount of temporal variation in the feature pdf. This figure illustrates the importance of using the right classifier architecture that takes advantage of the time-varying feature pdf.

5.3 Feature Rank Order Curves

Rank order curves are quite useful in determining an appropriate feature dimension for classification. Initially, as we add good features, the recognition performance increases. It reaches a plateau after a while and may even degrade as we increase the feature dimension beyond what is necessary. In short, the rank order curves are a useful tool to detect the occurrence of underfitting (i.e., using less features than necessary) or overfitting (i.e., using too many features). Figure 9 illustrates the rank order curve for the LFA target characterization.

5.4 Recognition Performance Comparison

We initially train the three classifiers with clean data. Next we test them with noise corrupted data. Figure 10 illustrates their recognition performance as a function of noise standard deviation.

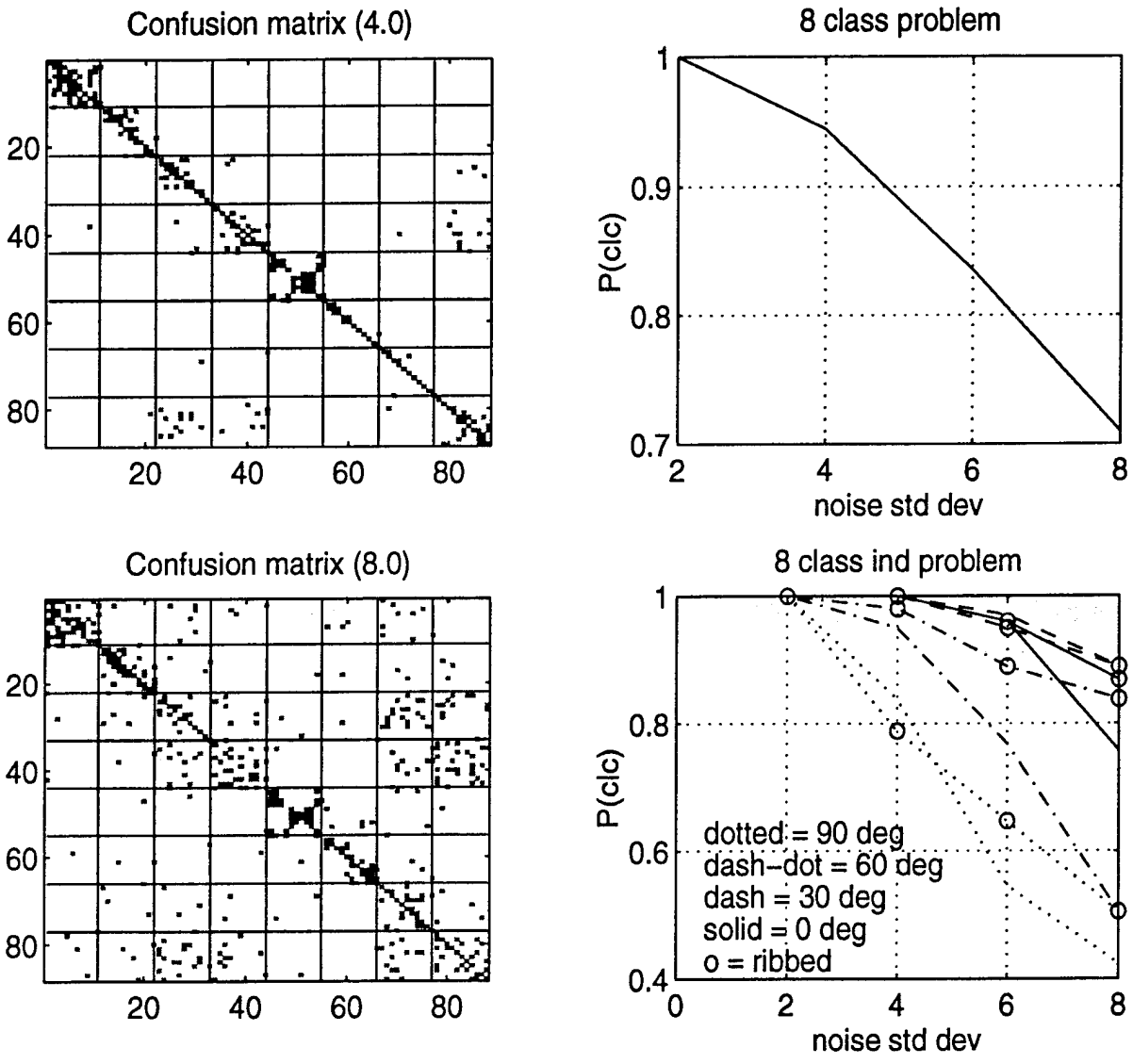


Figure 7: Recognition performance achieved by using the received waveforms as matched filters for aspect and target type variation. The 8-by-8 confusion matrices show the degree of confusion with other classes. The numbers in the parenthesis represent the noise standard deviation. The order of classes from top to bottom and left to right is as follows: empty0, empty30, empty60, empty90, ribbed0, ribbed30, ribbed60, and ribbed90. The y-axis represents the true class while the x-axis shows the classifier output.

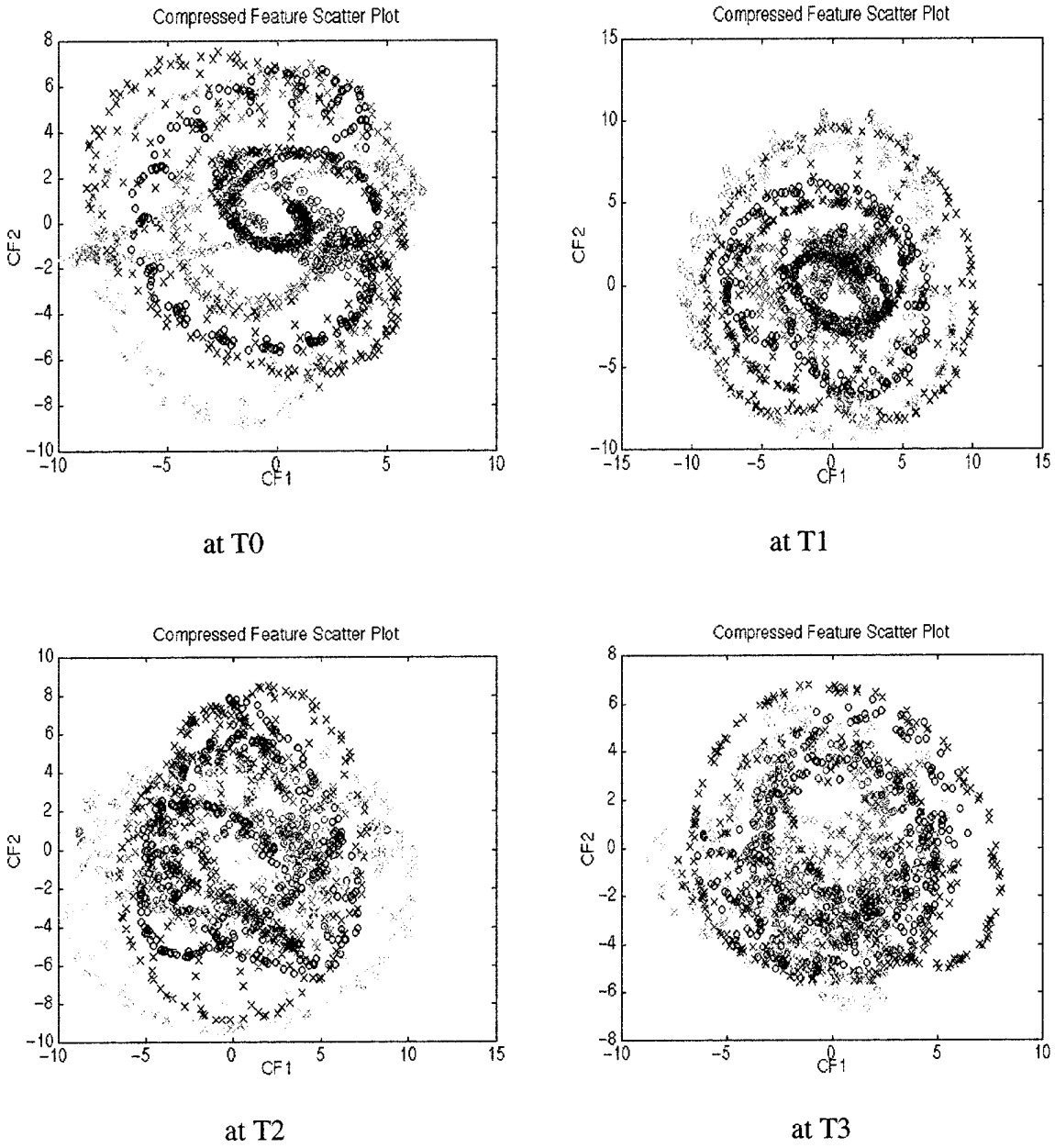


Figure 8: Compressed feature pdf plots show a highly non-Gaussian, multi-modal, and time-varying structure.

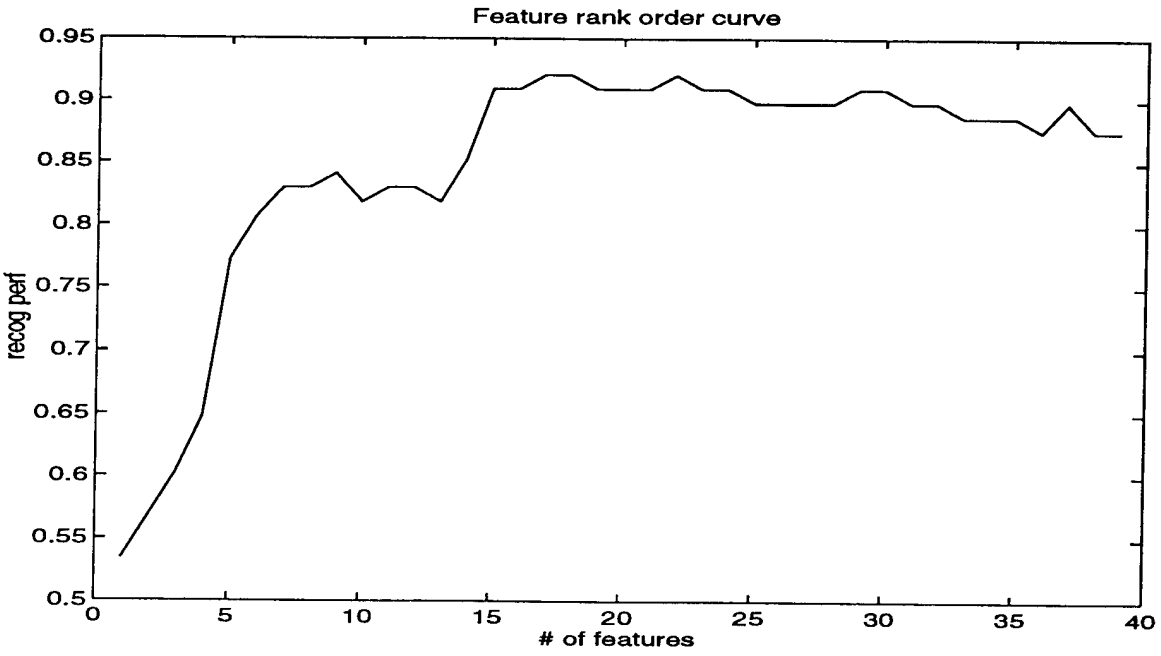


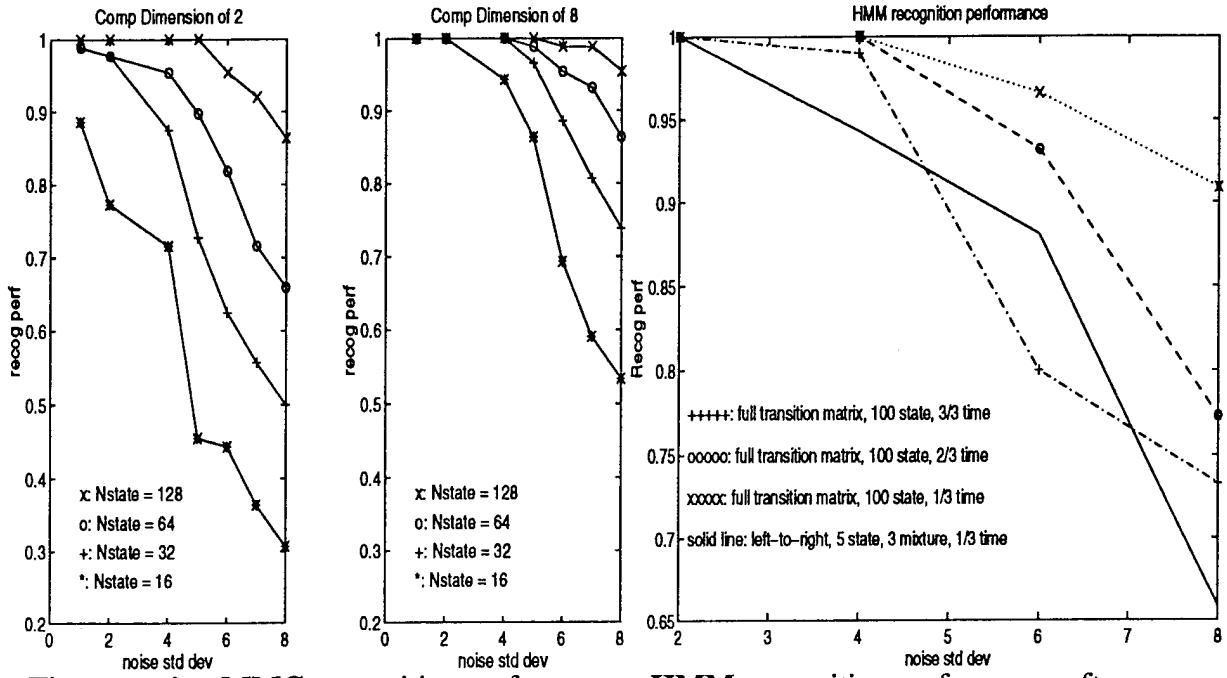
Figure 9: Rank order curves are useful in determining an appropriate number of features for maximum and robust recognition performance.

5.5 Discussion

It is interesting to note that the three classifiers outperform the matched filter-based recognizer, especially at low SNR. Initially, we expected the matched filter to provide the theoretical upper bounds (i.e., similar to the Cramer-Rao lower bounds on parameter estimation) on the LFA target characterization performance. Although this result appears to be contradictory at first glance, it makes sense in the context of classification. That is, if some temporal segments do not provide good inter-class separability, it is beneficial to remove those segments from classification. For detection, such a strategy will yield suboptimal detection performance.

Furthermore, it is worthwhile to investigate as many pertinent projection spaces as possible, provided that each projection space yield orthogonal features. For instance, we achieved 69.3 %, 76.1 %, and 46.6 % recognition performance from RID, matched filter (transmit waveform), and PCI-derived features, respectively. When we fused all the good features, our recognition performance improved to 92.1 %. Due to the broad-band nature of the transmit pulse, the PCI-derived features do not work as well as those derived from the other projection spaces. PCI is more appropriate when the transmit waveform exhibits a time-varying narrow-band structure (i.e, LFM, HFM, or FSK).

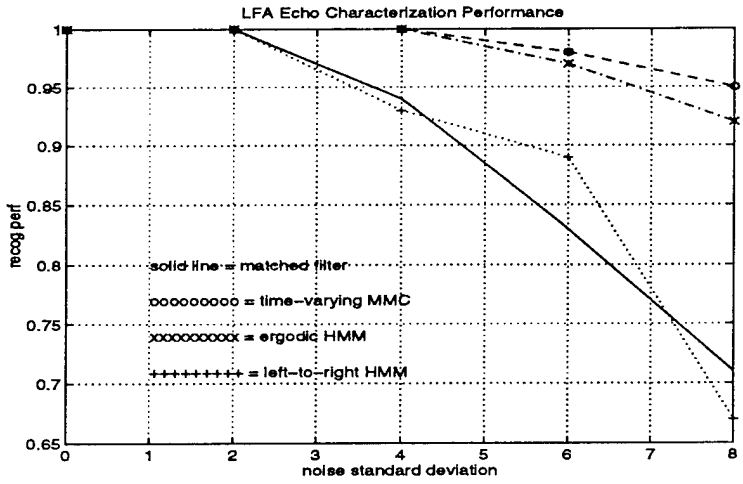
In order to demonstrate the importance of matching the classifier architecture to the underlying feature pdfs, we replace the multi-modal classifier (MMC) with the MVG classifier in the temporally adaptive classifier topology. Figure 11 illustrates a dramatic performance difference between the two classifiers. The left plots show the individual and overall target characterization performance as a function of time.



Time-varying MMC recognition performance with a different number of features as a function of the number of states and noise standard deviation.

HMM recognition performance after feature optimization as a function of the time segments and noise standard deviation.

Performance Summary of the Above



The key to achieving robust target characterization performance:
(a) energy compaction via low dimensional projection,
(b) thorough feature optimization,
(c) matching the classifier topology,
(d) and temporally adaptive LLR score integration.

Figure 10: Target characterization performance as a function of noise standard deviation, classifier topology, and feature.

The y-axis represents the recognition performance with an offset of 1.0 added for each case. That is, the overall recognition performance with the "+" legend should read 0.0 to 1.0 because of the offset of 4.0.

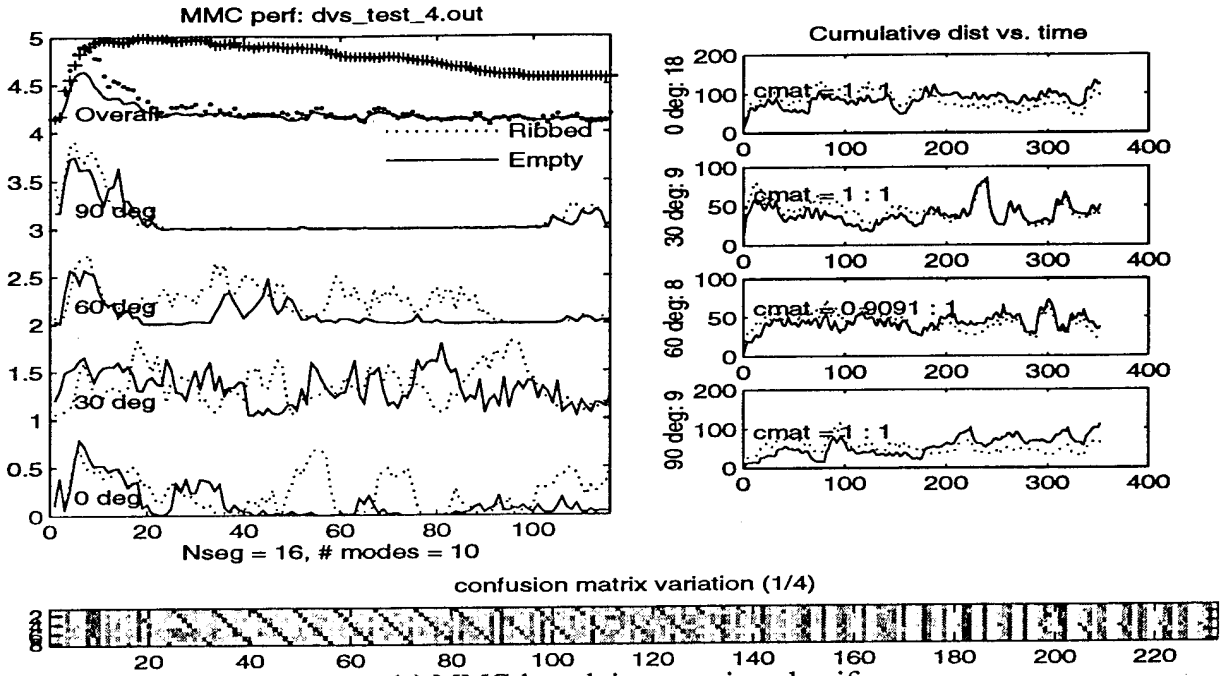
6 CONCLUSION & FUTURE DIRECTION

In this paper, we demonstrated the crucial link between target physics and signal processing with the scaled model data for the two cylinder types at the four aspect groups. We also developed the three recognizer topologies that exploited the underlying time-varying feature distributions. We also discussed LFA target echo formation process and the integrated classification paradigm that exploits an inherent relationship between features and classifiers. In short, a combination of (1) energy compaction via low-dimensional projection, (2) thorough feature discriminant analysis as a function of time, and (3) appropriate classifier topology generation is a key to achieving robust active target characterization performance.

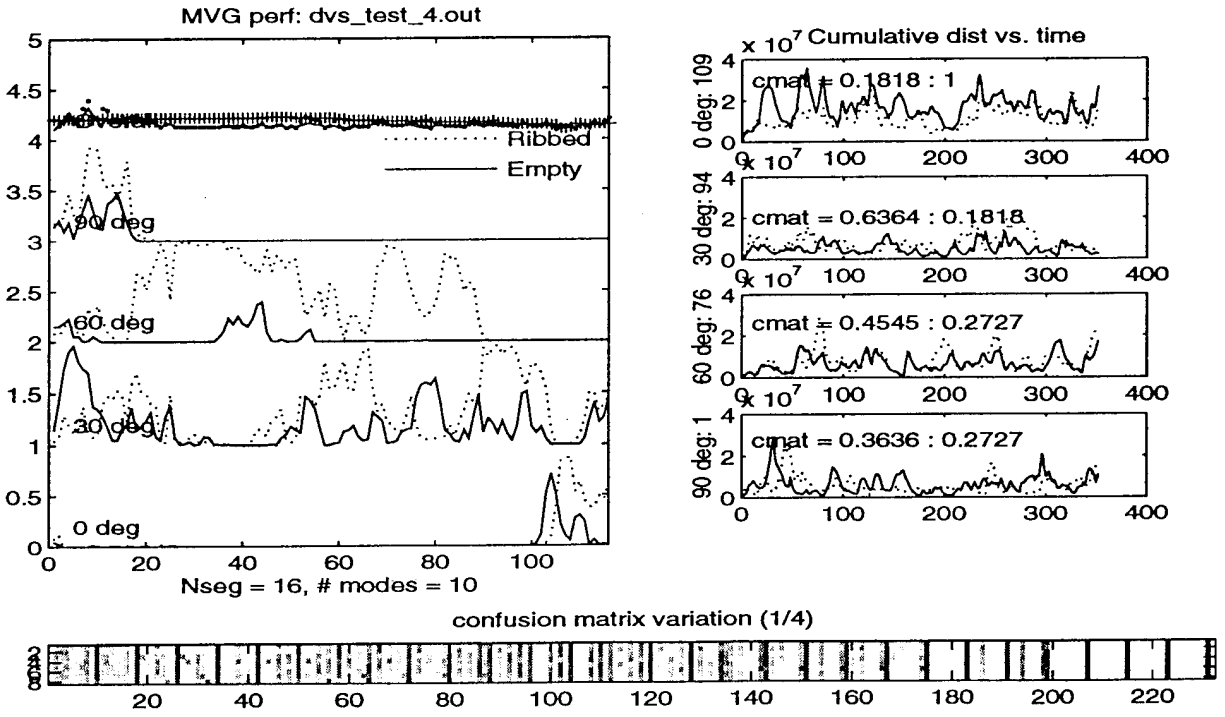
With the noise standard deviation of 6, the matched filter-based classifier yields 83.5 % correct recognition performance of all eight classes while the time-varying MMC, left-to-right HMM, and ergodic HMM are able to achieve 98 %, 89 %, and 97 % correct echo characterization, respectively. This improved performance is attributable to feature optimization and selection of an appropriate classifier topology for this problem. Furthermore, fusion of features derived from RID, segmented matched filter, and PCI projection spaces results in 16 % improvement in target characterization performance.

Since we demonstrated an excellent target characterization performance with the scaled model tank data, the next natural extension is for more realistic targets. The first area of future research is to apply the same target characterization algorithms to real world threats, such as scaled model submarines and mines. We can characterize target recognition performance in terms of confusion matrices and classification receiver operating characteristics (ROC) curves as a function of frequency, bandwidth, aspect, and time.

Furthermore, it is crucial that we investigate the impacts of confusion factors, such as ambient noise, clutter, and rapidly fluctuating channel responses in shallow water, on the target characterization performance. We are currently looking into blind deconvolution, hypothesis-directed matched field processing for depth and channel estimation, and channel deconvolution using probe pulses as a potential means of deconvolving the medium effects out of the received waveform. The key to successful target characterization is to remove as much confusion as possible prior to feature extraction and classification.



(a) MMC-based time-varying classifier



(b) MVG-based time-varying classifier

Figure 11: This example illustrates the importance of using the appropriate time segments that offer good class discriminability and the right classifier architecture that matches the underlying feature pdf. At low SNR, the late returns are highly corrupted by noise, thereby rendering them less than useful for discrimination. Furthermore, since the features exhibit non-Gaussian and multi-modal characteristics, MMC outperforms MVG by a wide margin.

7 ACKNOWLEDGEMENTS

This work was supported in part by the Office of Naval Research under contract N00014-91-C-0125. The technical reviewer for this program was Dr. Kam Ng of ONR, Code 122.

References

[1] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE*, vol.77, no.2, February 1989.

[2] K.-Y. Su and C.-H. Lee, "Speech Recognition Using Weighted HMM and Subspace Projection Approaches", *IEEE Trans. on Speech and Audio Processing*, Vol.2, No.1, Part 1, January 1994.

[3] Charles N. Corrado Jr., Mid-Frequency Acoustic Backscattering from Finite Cylindrical Shells and the Influence of Helical Membrane Waves, Ph.D. Thesis, Massachusetts Institute of Technology (1993).

[4] J. Jeong and W.J. Williams, "Kernel Design for Reduced Interference Distributions", *IEEE Trans. on Signal Processing*, Vol.40, No.2, February 1992.

[5] Joseph E. Bondaryk, Array Processing and Forward Modeling Methods for the Analysis of Stiffened, Fluid-Loaded Cylindrical Shells, Ph.D. Thesis, Massachusetts Institute of Technology (1994).

[6] Matthew Conti, Y. P. Guo, and Ira Dyer, "Influence of endcaps on mid-frequency scattering from finite cylinder shells at axial incidence", *J. Acoust. Soc. Am.*, 93, 2413 (1993).

[7] M. Wax and I. Ziskind, "Detection of the Number of Coherent Signals by the MDL Principle", *IEEE Trans. on ASSP*, Vol. 37, No. 8, August 1989.

[8] D.W. Tufts, D.H. Kil, and R.R. Slater, "Reverberation Suppression and Modeling", Ocean Reverberation edited by D.D. Ellis, J.R. Preston, and H.G. Urban, *Kluwer Academic Publishers*, Boston, MA, 1993.

[9] C.S. Myers, A. Singer, F. Shin, and E. Church, "Modeling Chaotic Systems with Hidden Markov Models", *Proceedings of ICAASP*, 1992.

[10] G.H. Golub and C.E. Van Loan, Matrix Computations, *Johns Hopkins Press*, Baltimore, MD, 1989.

[11] T.E. Luginbuhl, S.G. Greineder, and R.L. Streit, "Feature Set Reduction Algorithms for Probabilistic Neural Networks", *U.S. Navy Journal of Underwater Acoustics*, Vol. 43, No.2, April 1993.

[12] P.A. Devijver and J. Kittler, Pattern Recognition: A Statistical Approach, *London: Prentice-Hall*, 1982.

[13] A.J. Viterbi, "Error Bounds for Convolutional Codes and Asymptotically Optimum Decoding Algorithm", *IEEE Trans. on Information Theory*, Vol. IT-13, April 1967.

[14] D.H. Kil and F.B. Shin, "Reduced Dimension Image Compression for Remotely Distributed Underwater Signal Processing", submitted to the *1995 International Conference on Image Processing*, January 1995.

OFFICE OF THE UNDER SECRETARY OF DEFENSE (ACQUISITION)
DEFENSE TECHNICAL INFORMATION CENTER
CAMERON STATION
ALEXANDRIA, VIRGINIA 22304-6145

IN REPLY REFER TO DTIC-OCC

SUBJECT: Distribution Statements on Technical Documents

TO: OFFICE OF NAVAL RESEARCH
CORPORATE PROGRAMS DIVISION
ONR 353
800 NORTH QUINCY STREET
ARLINGTON, VA 22217-5660

1995 1031 040

1. Reference: DoD Directive 5230.24, Distribution Statements on Technical Documents, 18 Mar 87.

2. The Defense Technical Information Center received the enclosed report (referenced below) which is not marked in accordance with the above reference.

FINAL REPORT
N00014-91-C-0125
TITLE: LOW-FREQUENCY ACTIVE TARGET CHARACTERIZATION USING HIDDEN MARKOV MODELS AND CLASSIFIERS

3. We request the appropriate distribution statement be assigned and the report returned to DTIC within 5 working days.

4. Approved distribution statements are listed on the reverse of this letter. If you have any questions regarding these statements, call DTIC's Cataloging Branch, (703) 274-6837.

FOR THE ADMINISTRATOR:

1 Encl

GOPALAKRISHNAN NAIR
Chief, Cataloging Branch

FL-171
Jul 93

DISTRIBUTION STATEMENT A:

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

DISTRIBUTION STATEMENT B:

DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES ONLY; (Indicate Reason and Date Below). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT C:

DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES AND THEIR CONTRACTORS; (Indicate Reason and Date Below). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT D:

DISTRIBUTION AUTHORIZED TO DOD AND U.S. DOD CONTRACTORS ONLY; (Indicate Reason and Date Below). OTHER REQUESTS SHALL BE REFERRED TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT E:

DISTRIBUTION AUTHORIZED TO DOD COMPONENTS ONLY; (Indicate Reason and Date Below). OTHER REQUESTS SHALL BE REFERRED TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT F:

FURTHER DISSEMINATION ONLY AS DIRECTED BY (Indicate Controlling DoD Office and Date Below) or HIGHER DOD AUTHORITY.

DISTRIBUTION STATEMENT X:

DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES AND PRIVATE INDIVIDUALS OR ENTERPRISES ELIGIBLE TO OBTAIN EXPORT-CONTROLLED TECHNICAL DATA IN ACCORDANCE WITH DOD DIRECTIVE 5230.25, WITHHOLDING OF UNCLASSIFIED TECHNICAL DATA FROM PUBLIC DISCLOSURE, 6 Nov 1984 (Indicate date of determination). CONTROLLING DOD OFFICE IS (Indicate Controlling DoD Office).

The cited documents has been reviewed by competent authority and the following distribution statement is hereby authorized.

A

(Statement)

OFFICE OF NAVAL RESEARCH
CORPORATE PROGRAMS DIVISION
ONR 353
800 NORTH QUINCY STREET
ARLINGTON, VA 22217-5660

(Controlling DoD Office Name)

(Reason)

(Controlling DoD Office Address, City, State, Zip)

DEBRA T. HUGHES
DEPUTY DIRECTOR
CORPORATE PROGRAMS OFFICE

(Signature & Typed Name)

(Assigning Office)

19 SEP 1995

(Date Statement Assigned)